DEFENSE EQUAL OPPORTUNITY MANAGEMENT INSTITUTE

DIRECTORATE OF RESEARCH

Measurement and Construct Equivalence
of
Three MEOCS Scales Across
Eight Sociocultural Groups

by

Robert M. McIntyre, Ph.D.
Department of Psychology
Old Dominion University

Summer 2000

20010418 102

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
28-02-2001	Research	May - August 2000

4. TITLE AND SUBTITLE

Measurement and Construct Equivalence of Three MEOCS Scales Across Eight Sociocultural Groups

5a. CONTRACT NUMBER

N00014-97-1-1005

5b. GRANT NUMBER

N/A

5c. PROGRAM ELEMENT NUMBER

N/A

6. AUTHOR(S)

Robert M. McIntyre

5d. PROJECT NUMBER

N/A

5e. TASK NUMBER

N/A

5f. WORK UNIT NUMBER

N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Directorate of Research
Defense Equal Opportunity Management Institute
740 O'Malley Road MS9121
Patrick AFB, FL 32925-3399

8. PERFORMING ORGANIZATION REPORT NUMBER

RSP 00-06

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22302

10. SPONSOR/MONITOR'S ACRONYM(S)

ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

N/A

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Three scales of the Military Equal Opportunity Climate Survey (MEOCS) were examined for measurement and construct equivalence across eight sociocultural groups - White and African-American enlisted and officers of both genders. Two random samples of 1000 were drawn for each group. Weighted least squares and maximum likelihood procedures were used to assess factorial invariance, as well as invariance of indicator residuals, variances of the latent constructs, covariance of the latent constructs, and a subset of mean structures across the eight groups. Results indicated that measurement equivalence existed across the groups, which allowed for examination of construct equivalence. In the construct equivalence analyses, differences were found on mean structures between the enlisted African-American men and other groups. Nonequivalence was also found across the groups on the variances and covariances of the latent constructs. Discussion focused on the practical as well as theoretical implications of the differences that were found. In addition, recommendations were presented for using structural equation modeling in future examination of the various attitude scales administered by the Department of Defense.

15. SUBJECT TERMS

Military Equal Opportunity Survey, MEOCS, Measurement Equivalence, Construct Equivalence

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			J. C. Scarpate
U	U	U	UU	42	19b. TELEPHONE NUMBER *(Include area code)* (321) 494-2676

Measurement and Construct Equivalence
of
Three MEOCS Scales Across Eight Sociocultural Groups

Robert M. McIntyre, Ph.D.
Associate Professor of Psychology
Old Dominion University

ABSTRACT

Three scales of the Military Equal Opportunity Climate Survey (MEOCS) were examined for measurement and construct equivalence across eight sociocultural groups—White and African-American enlisted and officers of both genders. Two random samples of 1000 were drawn for each group. Weighted least squares and maximum likelihood procedures were used to assess factorial invariance, as well as invariance of indicator residuals, variances of the latent constructs, covariance of the latent constructs, and a subset of mean structures across the eight groups. Results indicated that measurement equivalence existed across the groups, which allowed for examination of construct equivalence. In the construct equivalence analyses, differences were found on mean structures between the enlisted African-American men and other groups. Nonequivalence was also found across the groups on the variances and covariances of the latent constructs. Discussion focused on the practical as well as theoretical implications of the differences that were found. In addition, recommendations were presented for using structural equation modeling in future examination of the various attitude scales administered by the Department of Defense.

Summer 2000

Measurement and Construct Equivalence
of
Three MEOCS Scales Across Eight Sociocultural Groups

Robert M. McIntyre, Ph.D.
Assistant Professor of Psychology
Old Dominion University

This report has three purposes: First, the concepts and importance of measurement invariance and construct invariance are described. Second, three scales comprising the Military Equal Opportunity Climate Survey (MEOCS) were analyzed to establish the degree of measurement and construct invariance across eight groups. Third, recommendations for future scale development and research are presented.

Background

McIntyre (1999) used structural equation modeling (SEM) in an effort "to confirm" the MEOCS' measurement structure that had been established by multiple exploratory factor analyses. This section on the background of the MEOCS borrows from the McIntyre 1999 report. The interested reader is encouraged to review the complete document.

Dansby and Landis (1991) defined equal opportunity climate as follows:
> ...The expectation by individuals that opportunities, responsibilities, and rewards will be accorded on the basis of a person's abilities, efforts, and contributions, and not on race, color, sex, religion, or national origin. It is to be emphasized that this definition involves the individual's perceptions and may or may not be based on the actual witnessing of behavior (p. 392).

According to this definition, the MEOCS assesses individuals' attitudes (perceptions, feelings, and beliefs) pertaining to equal opportunity (EO) fairness in the workplace. In addition, the MEOCS was originally designed to assess important organizational outcomes (e.g., job satisfaction, commitment to work, and perceived work group effectiveness) in order to understand the relationship between these outcomes and EO fairness attitudes. Over the years, the Directorate of Research at the Defense Equal Opportunity Management Institute (DEOMI) has carried out separate statistical analyses with hundreds of thousands of respondents[1] and has replicated the initial principal components analyses that was used to identify the original factors or scales. (For example, see Dansby and Landis, 1991). Results consistently have found the following set of factors:

---

[1] In fact, the current MEOCS data base comprises more than 1,000,000 responses

Table 1. Original Scales and Descriptions

Scale Designation	Scale Full Name	Type of Scale
DCBTM	Differential command behavior toward minorities	Fairness
PEOB	Positive Equal Opportunity Behaviors	Fairness
R_SB	Racism-Sexism	Fairness
RD	Reverse Discrimination	Fairness
SH_D	Sexual Harassment and Discrimination	Fairness
RD2	Reverse Discrimination II	Fairness
SAT	Job Satisfaction	Outcome
EFF	Perceived Work Group Effectiveness	Outcome
COM	Commitment	Outcome
DTMW	Discrimination toward minorities and women	Fairness
R_Gsep	Racial-Gender Separatism	Fairness
OEOC	Overall EO Climate	Fairness

In sum, the MEOCS is a multifaceted measure of attitudes based on more than a decade of empirical work. The MEOCS represents a high-quality measurement instrument based on (a) its empirical beginnings, (b) the technical care taken in its development, (c) its high internal consistency within scales (all above .75), and (d) its wide acceptance by the Services.

As indicated above, McIntyre (1999) examined the MEOCS factor structure to determine whether the scale or factor structure could be empirically confirmed through the application of structural equation modeling (SEM). In addition to carrying out a confirmatory factor analysis (CFA), McIntyre attempted to examine the MEOCS to determine if the instrument's scales might be improved. To this end, he modified the MEOCS scales primarily by removing scale items with low item reliabilities, high cross loadings with factors other than the factor to which they had been assigned, or high degrees of correlated measurement error. In order to accomplish these goals, McIntyre engaged in a three-stage study and at each stage examined the fit of the model to determine the effect of scale modification on measurement quality.[2] The final result of the study indicated that the MEOCS factor structure at the factor (scale) level was largely supported. All scales except the Reverse Discrimination II survived the scale trimming process. Further, the results cross-validated on a separate sample of 15,000. Nevertheless, several of the surviving scales were highly correlated (See McIntyre, 1999).

---

[2] The only alteration that McIntyre deemed appropriate in the structural equation modeling that he carried out was item deletion. Although he had the option of allowing for cross-loading items, or correlated measurement error-- which may have improved the various fit indices, these options would have resulted in scales with degraded measurement properties.

How MEOCS Data Are Used in the Field

     The MEOCS is intended as a tool that military commanders can use to assess the EO-related attitudes within their units and to take action when necessary. Dansby (1993) describes a six-step model that is followed in administering the MEOCS.[3]

     Step 1--Contact. The unit commander's representative contacts DEOMI's Directorate of Research (DR) to have MEOCS administered. Contact takes place, usually, as a result of the positive reputation that the MEOCS has had over the years. Requiring that the MEOCS be requested by unit commanders ensures that commanders are taking responsibility for its administration and for dealing with the derived information.

     Step 2--Contract. Once contacted, DR takes the initiative to explain what the MEOCS can and cannot provide. This is the step in which a psychological contract is struck between the requesting unit and DR. The responsibilities of the unit and its commander are laid out. Any special requests by the unit commander are addressed as well.

     Step 3--Data Gathering. The requesting unit's project officer (PO) is responsible for data gathering. DR communicates with the PO to ensure that he or she understands that all data are confidential, and should be collected in a timely manner to ensure the highest quality assessment.

     Step 4--Data Analysis. DR analyzes the data returned by the PO. Data are analyzed in an automated three-stage process including optical scanning of response sheets, statistical analyses of the scanned data, and compilation of the data into a databased report. The report represents the feedback that is sent to the unit commander for his or her examination. The information summarized in the report includes:

    a. An executive summary containing introductory explanations, frequency reports on the number of responses by sociocultural subgroup, comparisons of the overall unit factor score means with the DEOMI database means for all services and the appropriate service, and graphic as well as numeric representations for the overall comparisons and major subgroup comparisons.

    b. Subgroup comparisons that are statistically significant are presented as "statistically reliable differences."

    c. A disparity index is presented which represents the overall disparity in viewpoint between subgroups. The goal is to identify possible problem areas in a succinct manner.

     Step 5--Feedback. As indicated above, the feedback consists of an executive summary designed to be easily readable by busy unit commanders. Besides the data described above, recommendations for follow-up interventions are made. The complete feedback report includes not only an executive summary containing most of what a unit commander would need, but in-depth explanations of the philosophical underpinnings of the survey and the data summarized therein.

     Step 6--Follow-up. Many unit commanders will choose to follow-up the collection of data with certain actions that might alleviate newly discovered problem areas. Recommendations for follow-up are presented in the MEOCS feedback package. However, if unit commanders would like to pursue different or additional follow-up interventions, they are encouraged to discuss these with DR.

     This short review of the steps followed in administering the MEOCS is directly pertinent to the second purpose for carrying out the present research. As stated above, the current study

---

[3] Note: Dansby points out that these six steps parallel the classical approaches to organizational development and survey feedback described in the organizational science literature.

examines the degree to which the measures comprising the MEOCS are equivalent across groups. Unless measures are equivalent, then unit commanders may draw incorrect conclusions from the feedback or may have difficulty making sense of the feedback.

Measurement and Construct Equivalence[4]

In discussing the comparability of measures across groups, there are two important distinctions to be made: measurement equivalence and construct equivalence. (These terms are recommended by Little (1997)).

Measurement equivalence. In psychometrics, it has long been known—but sometimes ignored—that a necessary condition for examining mean differences between groups on constructs assessed by some measurement instrument (e.g., a test or attitude questionnaire) is that the items comprising the measure (the indicator variables) are phenomenologically equivalent across the groups (Alwin & Jackson, 1981; Thurstone, 1935; Thurstone, 1947). Phenomenological equivalence is more commonly referred to as measurement equivalence or factorial invariance. These terms imply that groups of individuals (such as sociocultural groups) attribute similar meaning to the constituent items of scales comprising the measurement device.

Technically, measurement equivalence has several statistically demonstrable characteristics. First, the same number of latent variables with the same pattern of loadings should fit data from different groups. Second, the relative importance of the items comprising the scales must be equal.[5] "Relative importance" is represented by factor loadings from a factor analysis. Factor loadings are defined as the regression coefficients computed when an item— i.e., an indicator of a factor—is regressed on a factor or latent construct. Most experts in structural equation modeling (SEM) deem these two conditions as necessary and sufficient for follow-up comparisons of means on the measurement instruments (e.g., Byrne, 1998 and Kline, 1998)[6]. However, Little (1997; 2000), in explaining Mean and Covariance Structures (MACS) Analysis, indicates that the intercepts or means of the indicator variables (the items) should be tested for equivalence across groups. This requirement is not specifically discussed by many researchers in the field of psychological measurement (e.g., Byrne, 1998; Kline, 1998; Jöreskog & Sörbom, 1996). There is one final characteristic that is sometimes examined to establish measurement equivalence—that the residual variances (unique factors and unreliability) of the items are equal. However, it is generally thought that such measurement equivalence at this level is far too stringent a condition for legitimizing the comparison of means across groups.

Construct equivalence. Once measurement equivalence is established, construct equivalence can be examined. Construct equivalence is a term that Little (1997) used to describe comparisons of groups at the construct level. One such comparison has already been mentioned: comparison of means of the latent constructs. In addition, construct equivalence may involve a comparison of latent variances and comparison of covariances among the latent constructs.

---

[4] The terms measurement equivalence and measurement invariance are used interchangeably in this report. Construct equivalence and construct invariance are also used interchangeably.

[5] Technically, the relative importance (defined as factor loadings—see below) must be demonstrated to be not statistically significantly different.
[6] Technically, total equality on factor loadings is not required. For example, Byrne (1998) explains that **partial** measurement invariance in which the majority (rather than all) of the factor loadings are fixed as equal across groups is necessary for comparing means across groups.

Theoretical and Practical Importance of Both Types of Invariance

The MEOCS is designed primarily for the practical purpose of organizational development. As summarized above, unit commanders request an administration of the MEOCS in order to assess the state of equal opportunity climate within their units. The feedback that unit commanders receive involves, among other things, comparisons of scale means of the various MEOCS factors (latent constructs). As stated above, psychometric research indicates that in order to make such comparisons meaningful, one must examine the measurement equivalence of the MEOCS scales. This points to the most important practical reason for assessing measurement equivalence.

There are theoretical reasons as well to examine measurement equivalence. For example, research has been carried out over the years examining (a) differences between sociocultural groups comprising the MEOCS database and (b) comparison of relationships among MEOCS latent constructs and other ancillary data imbedded in the MEOCS database. Both of these research domains first require measurement invariance of the constructs involved in analysis.

Are there practical reasons for examining construct equivalence? Clearly, mean comparisons presented to unit commanders directly answer the question. But there are other comparisons as well that may have practical implications. For example, some have recommended investigating the difference in correlational structure among latent constructs across different sociocultural groups. Taris, Bok, and Meijer (1998) indicated that differences across sociocultural groups (e.g., gender, race, ethnic, status groups) can take on three forms analogous to the three forms of intervention-caused change that Mortimer, Finch, and Kumka (1982) and Golembiewski, Billingsley, and Yeager (1976) discuss. These three forms are alpha change (i.e., differences in level on some construct), beta change (i.e., recalibration of self-perceptions on some construct), and gamma change (i.e., a reconceptualization of the interrelationships among the facets of some multifaceted construct). Most apropos of comparisons across sociocultural groups is the gamma change or, more precisely in the context of comparing sociocultural groups, **gamma difference**. Gamma difference addresses the issue of whether the structure of some multifaceted construct (such as job satisfaction, perceived fairness, and job commitment) is different across different groups. Having both practical as well as theoretical implications, gamma difference will lead to a much richer understanding of different sociocultural groups and may well lead to far more effective training and organizational development interventions.

McIntyre (1997) examined the response patterns and differences in variance of responses to MEOCS. He found that different sociocultural groups showed statistically significant differences in variances. This research suggested that there might be a need for attending to variance differences in addition to mean differences across sociocultural groups in order to optimize training or other organizational development interventions.

5

Establishing Measurement and Construct Equivalence

Table 2 (see McIntyre (1999) presents a set of common terms for the reader who may be only moderately familiar with SEM.

SEM provides the set of tools necessary for analyzing measurement and construct equivalence. The following are the steps involved in the process:

Examine the baseline models. The first step in examining measurement and construct equivalence across groups involves the investigation of the measurement model for each group individually. The heart of the measurement model encompasses the items' loadings on their respective latent constructs. Beyond this, the residual variances of each indicator variable constituting the measurement error and unique factor associated with the indicator may be examined. Correlations among residuals (often referred to as correlations of measurement error) also represents a part of the measurement model. Finally, if a construct is multifaceted—that is, there are multiple constructs comprising the latent construct domain—then the measurement model also involves the set of variances and covariances among the latent constructs.

Examining the baseline involves determining whether there is a reasonable fit of the data to the measurement model. Fit is assessed by means of the χ^2 value provided by LISREL (and other SEM programs). However, χ^2 are extremely sensitive to sample size, leading researchers to use what are referred to as "practical indices of fit." McIntyre (1999) provided a Table with a detailed summary of the common fit indices along with their generally preferred cutoff values. This table appears as Appendix 1 in the present document.

Three commonly used practical fit indices are the root mean square error of approximation (RMSEA), the non-normed fit index (NNFI), and the comparative fit index. As Appendix 1 indicates, RMSEA values less than .05 are considered very good fit. Values of .08 to .05 are considered acceptable fit. Values between .10 and .08 are considered mediocre fit. Values exceeding .10 are considered indication of poor fit. With regard to the NNFI and the CFI, values close to 1.0 are considered representing excellent fit and values below .90 are considered representative of poor fit. Most experts in measurement recommend that several fit indices be used in assessing the solution. A researcher examines the baseline model by first inspecting the fit indices. If there seems to be room for improvement, the researcher may carry out an analysis of the modification indices (MIs) which represent the difference in the overall model χ^2 if a particular parameter (such as a loading) were allowed to be free. For example, freeing a value may mean allowing an indicator to load on a factor other than the one that is was intended to indicate. It may also mean allowing measurement errors to covary. Large MI values point to parameters that might be freed in this way. Researchers must take into account the theoretical justification for freeing parameters. They must also be aware of the danger of taking advantage of chance in such manipulations of the measurement model.[7]

---

[7] It is important to recognize that any time a researcher modifies the measurement model by allowing an observed variable to load on a non-targeted latent variable or by allowing correlated measurement error after assuming no correlated measurement error, he or she has changed from a confirmatory analysis to an exploratory analysis. This is where cross-validating across multiple random samples from the same population will help reduce the likelihood of taking advantage of change.

Table 2
Common Terms Used in Structural Equation Modeling and Confirmatory Factor Analysis

Model	A term that refers to the causal structure among various observed and unobserved variables.
Model parameter	Any population characteristic estimated in a sample. The following are examples of model parameters: loadings, path coefficients, indicator reliabilities, covariances among latent constructs.
Observed variable	A variable for which data (measurements) exist.
Unobserved (latent) variable	A variable for which no measurement exists but which is hypothesized to exist. Latent variables are presumed to "cause" unobserved variables in confirmatory factor analysis. Another term for latent variable is factor or scale.
Measurement model	In structural equation modeling (SEM), the measurement model refers to the causal links between latent and observed variables.
Hypothesized model	The model that the researcher proposes to explain the measurement of latent and observed data.
Exogenous variable	A variable that causes another variable. Exogenous variables can be latent or observed.
Endogenous variable	A variable caused by another variable. Endogenous variables can be latent or observed.
Indicator variable	An observed item that is presumed to be caused by a latent variable. When an observed item is said to indicate the latent variable, it is also presumed to be caused by the latent variable.
Item reliability	The percent of variance in an item explained by the model. In the case of a model whose indicator variables (items) are presumed to load on a single factor, the item reliability is the squared correlation between the latent variable and the item.
A variable's error variance	A variable's variance unaccounted for the by model
Fit of a model	The degree to which the covariation among observed items is explained by the hypothesized model. There are many indices of model fit.
Nested Model	Let there be two structural equation models, A and B. B is nested in A if B can be created from A from imposing additional model constraints on A.
Trimming a model	When a model is altered after preliminary investigation of the CFA results, it is said to be "trimmed." Trimming involves freeing a parameter that was fixed or changing the assignment of an item to an additional or different factor.
Λ_x	P by K matrix of factor loadings where P is the number of indicator variables and K is the number of latent constructs or factors.
Φ	K by K variance-covariance matrix associated with the latent constructs
χ^2	Chi-square value assessing the fit of a solution

$\Sigma(\theta)$	This is the product of a vector (θ) that comprises the model parameters and the population covariance matrix (Σ). This is the model implied covariance matrix and is compared to the sample covariance matrix.
Σ	The population covariance matrix representing relationships among observable items in the population.
S	The sample covariance matrix representing the relationships among observable items in the sample.

Once the measurement model has passed muster, then the next step begins.

Simultaneous analysis of the pattern of loadings. SEM allows for a simultaneous assessment of the fit of a measurement model across multiple groups. In this step, the researcher places only one constraint on the model—that the same number of latent variables indicated by the same observed variables holds across all groups. If the fit indices suggest a poor fit, then the researcher has several alternatives. The first is to outright reject the hypothesis of invariance. The second is to modify the measurement model problem. This may entail focusing on a subset of the latent constructs that are expected to hold up across the groups or focusing on a subset of the groups.

Simultaneous analysis of the magnitude of loadings. In this step, the researcher places an additional constraint on the model—that all indicators presumed to be caused by the latent factors have the same loadings across groups. Recall that a latent variable loading is a regression coefficient (slope) where the observed variable is regressed on the latent variable. The constraint embodied in this step amounts to holding constant the value of all loadings across all groups.

The measurement model in this step is said to be nested in the model associated with the first step. Hayduk (1987) explains that a model is nested in another if imposing additional constraints on the other (original) model can create it. One way of imposing such constraints is to fix certain parameters of the original model to zero. Another approach is to constrain certain parameters to be equal across groups. The feature of nested models allows for a statistical comparison of the models. The difference between the χ^2 values from two nested models is itself distributed as a χ^2 with df equal to the difference between the subsuming and nested models. The difference in χ^2 is often referred to as $\Delta\chi^2$. If this value exceeds the value associated with a properly selected cut-off value (which takes into account control for Type I and Type II error rate), then one can conclude that the nested model is statistically inferior to the subsuming model. Unfortunately, χ^2 and $\Delta\chi^2$ are highly sensitive to sample size and to even slight deterioration in model fit. Therefore, most experts in SEM (Byrne, 1998; Jöreskog & Sörbom, 1996) caution users of SEM about the sole use of the χ^2 and $\Delta\chi^2$ measures of fit without reference to the so-called practical measures such as RMSEA, NNFI, and CFI. In fact, Little (1997) recommends that in examining the measurement equivalence, the practical fit measures are sufficient while in analyzing construct equivalence, the statistical index (viz. $\Delta\chi^2$) should be used.

8

Therefore, in the current research, assessment of measurement equivalence is accomplished through the use of the practical fit indices. If these fit indices show an acceptable fit, then one can assume that the latent variables are equivalent in meaning across groups. If the fit measures suggest a less than adequate fit, then it is possible to modify the original model to free up certain parameters within one or more of the groups. Once again, MIs can be examined for this purpose. While such tinkering with the solution has its drawbacks when there is no way of cross-validating the solution in a separately drawn random sample, Byrne (1998) suggests that the practice is acceptable and useful to science in the sense that other researchers are provided with guidance as to how their measurement models should be formulated. I agree with Byrne as long as the researcher provides a clear and complete explanation of the steps taken in the revision of the measurement model.

Byrne (1998) and Byrne, Shavelson, and Muthén (1989) (and others as well) suggest that if full invariance is not found in this stage, all is not lost. They hold—and most experts agree— that partial measurement invariance is sufficient to carry out construct equivalence analyses. Partial measurement invariance (equivalence) exists when the majority of the loadings in the Λ_x matrix (the matrix of factor loadings) are equal across the groups. In other words, in certain groups comprising the multiple-group analysis, a minority of factor loadings may be free to vary or free to "cross-load"—that is load on factors other than the target factor.[8]

Simultaneous analysis of indicator residuals. The residuals in the SEM framework refer to a composite of the errors of measurement and unique latent variables associated with the indicators. Most authors do not make a practical distinction between these two components because there is no way of isolating one from the other. It is often assumed that the residuals represent error variance in that they do not represent the common factor of interest. High values for residuals may imply that the indicator variables are reliable. In multi-sample SEM analyses, high values may be found for some variables (with associated high MIs) within some samples. This suggests different levels of indicator reliability in certain groups.

In many applications of confirmatory factor analysis and SEM, the measurement errors (indicator residuals) are assumed to be correlated zero among all indicators. If p is the number of indicators in the analysis, the p-by-p matrix, Θ_δ, contains the residual values on the diagonal while the off-diagonal elements are assumed to be zero. That is, Θ_δ is assumed to be a diagonal matrix. However, Θ_δ is not required to be diagonal. Nonzero off-diagonal elements represent correlations (technically, covariances) among measurement residuals or measurement error. SEM programs such as LISREL provide MIs for the off-diagonal elements. When such an MI has a high value, it signifies that there may be correlated measurement error between a pair of indicator variables. Off-diagonal elements can be allowed to take on nonzero values if there is theoretical justification for doing so.[9] Byrne (1998) discussed this as a viable option in

---

[8] Little (1997) holds that there is an additional constraint that must hold in order to justify construct equivalence analyses. That is, in addition to the constraint that a majority of the Λ_x values being equal across groups, the intercepts of the indicators (regressed on the latent variables) must be shown to be invariant. Few researchers discuss this position.

[9] Correlated measurement error may be due to the proximity of one questionnaire item to another in the questionnaire, the similarity of wording between one questionnaire item and another, and so on. It is possible that certain groups within a multiple-group analysis have more correlated measurement error on certain items than other groups.

improving the fit of the model. On the other hand, measurement purists may reject such a practice.

Most measurement experts agree that requiring the equivalence of indicator residual variances is an overly restrictive constraint. However, an analysis of equivalence of residual variances may be of theoretical interest. For example, it provides a means of determining the difference in indicator variable reliability across different sociocultural groups. It also may provide a means of isolating culturally determined effects on measurement at the item level.

Testing for the invariance of the Θ_δ matrix (containing parameters associated with indicator residual variances and covariances) is the last step in assessing measurement equivalence. The remaining steps involve assessing construct equivalence. The prerequisite for carrying out the following analyses is at least partial measurement invariance (that is, a majority of factor loadings are equal across all groups).

Simultaneous analysis of latent variances and covariances. In LISREL 8.3 terminology, ϕ is defined as a $\underline{k}$ by $\underline{k}$ matrix of variances and covariances of the latent constructs (where k is the number of latent constructs). SEM (LISREL) allows the researcher to test the invariance across groups for any or all elements in ϕ. In this regard, one might constrain to be equal the diagonal elements of ϕ (that is, the latent variances) and determine the $\Delta\chi^2$ reflects a degradation in fit. Note that Little (1997) suggests that questions pertaining to construct equivalence (such as the invariance of the latent variances) should be addressed via statistical tests (i.e., through $\Delta\chi^2$ values, adjusted for Type I and II errors) and not through the practical fit indices.

In similar manner, the covariances among the latent constructs can be tested for invariance. LISREL allows for constraining the off-diagonal elements of ϕ to be equal. Further, as was implied above, a single element in ϕ can be assessed through the nested model comparison process.

Simultaneous analysis of latent mean structures. The final aspect of construct equivalence concerns the means of the latent constructs. As Kline (1998) states,

> "The basic datum of SEM, the covariance, does not convey
> information about means. If only covariances are analyzed, then
> all observed variables are mean-deviated (centered) so that latent
> variables must have means of zeros...Means are estimated in SEM
> by adding what is known as a mean structure to the model's basic
> covariance structure..."

Byrne (1998) and Kline (1998) provide a review of some basic concepts in linear regression in order to understand the statement above. First, it must be stated that SEM assumes that an indicator variable (for example, an item in an attitude questionnaire) is regressed on a latent variable. It is typically the case that there are multiple indicator variables, each of which is regressed on a particular variable. Let's consider the case of one indicator variable (X) and one latent variable (Y). In the usual case of SEM, where covariances only are examined, the model describing is as follows:

$$X = \lambda Y + e \qquad (1)$$

where $\underline{X}$ is the indicator variable, $\underline{\lambda}$ is the regression coefficient, $\underline{Y}$ is the latent variable, and $\underline{e}$ represents the residual of $\underline{X}$ not accounted for by $\underline{Y}$. In mean structures analysis, the mean of $\underline{X}$ is essential for analyzing the mean of $\underline{Y}$. So the following linear regression equation takes the place of Equation 1:

$$X = \tau + \lambda Y + e \qquad (2)$$

where τ are the intercept and all other terms remain the same. If we take expectations on both sides of Equation 2, under the assumption that the expected value of $\underline{e}$ is zero, then we have Equation 3.

$$M_x = \tau + \lambda M_y \qquad (3)$$

This indicates that the means of the indicator variables are a weighted composite of the intercept plus the regression coefficient (in SEM, the indicator's loading) times the mean of the latent variable. Note that in order to estimate an intercept in simple linear regression, $\underline{Y}$ is actually an $\underline{N}$ by 2 matrix. Column 1 of this matrix is a vector containing $\underline{N}$ elements each equal to one. Column 2 contains the values of $\underline{Y}$.[10] The intercept (τ) is computed in Equation 3 by regressing $\underline{X}$ on the column vector of 1s.

SEM programs such as LISREL automatically create a constant (that is, the vector containing 1s) when the user requests means structures analysis. Hence, the intercepts associated with the observed indicator variables and their means can be structured in a way similar to that described in Equation 3. The mean of the latent variable is obtained by regressing the latent variable on the constant (vector of 1s). In LISREL parlance, the symbol for the mean of a latent variable is $\underline{\kappa}$. In actuality, $\underline{\kappa}$ is a vector on latent variable means. Therefore, Equation 4 is a rephrasing of Equation 3 to include this new term.

$$M_{xi} = \tau_{xi} + \lambda \kappa_{xi} \qquad (4)$$

One last addition to this brief overview of mean structures analysis is necessary. It is not possible for SEM to estimate the means of the latent variables in a single group analysis because such single-group analyses are underspecified (that is, they have too few data points relative to parameters to be estimated). Therefore, multiple-group analyses are usually required to investigate mean structures. Because the latent variable technically has no scale, the mean of one group, the reference group, is set to a value of zero. Thereafter, the latent means are presented as comparison between the reference group and one of the remaining groups. This fact is important because it means that technically it is not a simple process to compare any two groups on their means. Dickinson (personal communication, 2000) indicates that researchers must set up the LISREL analysis in such a way that they answer the primary questions. The point is that researchers must consider which of multiple groups is the logical reference group with which to make comparisons.

In analyzing mean structures, therefore, the researcher is interested in investigating whether the means of the latent constructs in several groups are different or similar to the latent means of a reference group. This is accomplished by fixing the $\underline{\kappa}$ values for the reference group to zero and freeing the $\underline{\kappa}$ values across the remaining groups.[11] Note that although reanalysis of the data can be done by changing the reference group, this results in statistically nonindependent comparisons.

---

[10] In ordinary (raw-score) linear regression, the two columns correspond to two parameters that are estimated. Hence, the degrees of freedom for simple linear regression is N (number of observations) minus 2.

[11] In mean structure analysis analyses, the intercepts of the observed variables as well as the loadings are assumed to be held constant across all groups.

One might wonder why someone would "go to the trouble of using SEM" when one could readily use multivariate analysis of variance (MANOVA) to accomplish a comparison of scale means across groups. There are two issues at least that lead one to use SEM over MANOVA. First, note that in SEM, error terms are explicitly estimated. This implies that SEM allows for a comparison of means holding constant measurement error, which in turn leads to a much more precise and appropriate analysis. Second, Cole, Maxwell, Arvey, and Salas (1993) explain that when a "variable system" is to be compared across groups, and this variable system consists of latent constructs for which the constituent variables (e.g., items on a scale) are indicators, then SEM is the appropriate strategy. Cole et al. (1993), referencing Bollen and Lennox (1991), point to a distinction between this type of variable system which logically requires SEM and another type referred to as an emergent variable system in which the system variables are in a sense caused by the constituent variables. From the argument presented by Cole et al., it would seem that SEM is usually the preferred method in comparing scales based on responses to attitude questionnaires or most psychological tests and measures.

Analytical Goals of the Present Study

The analytical goals of the present study are to test the measurement equivalence and the construct equivalence of three scales of the MEOCS. The three scales and their definitions are as follows:
1. Differential Command Behavior toward Minorities (DCBTM): The degree to which respondents perceive that minorities are treated different from majorities by military supervisors.
2. Discrimination toward women and minorities (DTWM): The degree to which respondents perceive that women and minorities are treated unfairly.
3. Racial-Gender Separation (RGSEP): The degree to which respondents believe that people of same race or gender should associate within their own respective groups.

These three MEOCS scales were selected for the following reasons. First, they cover a wide array of what might be referred to as EO-related fairness. Second, in comparison to several MEOCS scales, these faired well in the confirmatory factor analyses carried out by McIntyre (1999) in the sense that relatively few items were dropped on the basis of examining MIs and cross-loadings. Third, McIntyre (1999) found in his analyses that the latent variables corresponding to the EO-related scales (i.e., no commitment, satisfaction, or work group effectiveness) were highly correlated. The high level of correlations between the chosen three EO-related latent scales and the others (most above .85) suggested that three latent variables covered much of the content of the entire set of EO scales.

Specifically, eight groups were selected as targets for the analysis. These eight groups resulted from the intersectional analysis of gender, race (African Americans versus Whites), and military status (enlisted versus officer). These groups were selected for three reasons: First, the breakdown by African-American versus White represents an historically important comparison in the military and society in general. (Certainly, this is not to diminish the comparison of other ethnic groups to be carried out in future research.) Second, the breakdown by men versus women also represents an extremely important issue as more and more women enter the military. Third, the status of officers versus enlisted is an important one for understanding the tenor of EO attitudes within the military. The following multiple group analyses were carried out on eight groups:

1. Analysis of the constancy of the three-factor structure.
2. Analysis of the measurement equivalence.
3. Analysis of the error variance and covariance equivalence.
4. Analysis of the equivalence of the variances of the latent constructs.
5. Analysis of the equivalence of the covariances among the latent constructs.
6. Analysis of mean structures across groups.

Method

Population Data Base

In July 1999, the MEOCS database consisted of approximately 816,000 cases on 130 variables. In an effort to prepare this data base for sample selection, McIntyre (1999) first identified all cases with greater than 10 percent missing data on the MEOCS-specific items (that is, 100 items which excluded demographic information). Based on recommendations by Kline (1998), these data were dropped from the data set. For all cases remaining, which contained up to 10 percent missing data, the modal values from the entire data set were substituted for the missing values for each variable. The net population from which samples were drawn for the current student exceeded 588,000 cases.

Samples

Eight pairs of random samples corresponding to the eight groups to be compared were drawn from the modified MEOCS database. Each sample consisted of 1000 randomly selected observations. Two random samples per group were drawn in order to cross-validate any findings.

Analytic Goals

The following were the analytic goals in the current study. As a set, these goals represent a comprehensive strategy for assessing the measurement and construct invariance of any measurement instrument.

1. Analysis of the baseline models—that is, a confirmatory factor analysis of the three-factor structure in each of the eight groups. (Note: this represents a series of single-sample analyses. The remaining are multi-group analyses.)
2. Analysis of the constancy of the three-factor structure.
3. Analysis of the measurement equivalence—invariance of the factor loadings.
4. Analysis of the equivalence of the variances of the latent constructs.
5. Analysis of the equivalence of the covariances among the latent constructs.
6. Analysis of the equivalence of the mean structure across the eight groups.
7. Analysis of the error variance and covariance equivalence.

Analytic Approaches

Two analytic approaches were used in this study. The first involved the use of weighted least squares (WLS) procedures on polychoric correlation matrices. The second involved the use of maximum likelihood (ML) procedures on parcels (linear composites) of items comprising the scales.

WLS. PRELIS (from the LISREL program) was used to compute a polychoric correlation matrix between indicator (observed) variables and to compute an associated asymptotic covariance matrix. A polychoric correlation represents an estimated linear relationship between two polychotomous variables assumed to be continuous and bivariate normal. Jöreskog and Sörbom (1993) recommend that the polychoric matrix be used because most attitudinal scale measurement employs polychotomous measures (for example, items with response scales ranging in discrete values from 1 to 5). Technically, these variables are ordinal in nature but are usually assumed to represent underlying continuous measurements. In computing a polychoric correlation matrix, LISREL employs pairwise contingency tables of the ordinal variables. For an explanation of the issues pertaining to polychotomous data, Jöreskog and Sörbom (1993) recommend Jöreskog and Sörbom (1988), Jöreskog (1990), and Jöreskog and Aish (1994). LISREL analyzes Polychoric correlation matrices through the application of the WLS estimation procedure. The weight matrix required in WLS is the inverse of the asymptotic covariance matrix (ACM) which also is computed by PRELIS (Jöreskog & Sörbom, 1993).

The use of WLS analysis with polychoric correlations provides an accurate solution in the SEM analysis (Jöreskog & Sörbom, 1996) while the ML method requires continuous, multivariate normal data and is inappropriate with polychoric data. However, if ML is the preferred analytic method in cases where Likert-style questionnaire data are analyzed, *parcels* (small composites) of items comprising the measure should be created. Covariance matrices are then computed on the parcels and serve as input for ML analyses. At times, the use of parcels can present certain disadvantages. For example, a researcher may prefer to work at the item-level in assessing measurement instruments to identify "problem items" that may account for an inadequate fit of a model. In the present study, both WLS and ML were used.

Unfortunately, WLS analyses involving polychoric matrices at present cannot be used for examining mean structures because it operates on a correlation matrix (Gerhard Mels, a technical advisor from Scientific Software International, the organization that markets LISREL, July 29, 2000, personal communication). A correlation matrix, in effect, is a covariance matrix computed on standardized data. Because standardized data are defined to have means of zero, they provide insufficient information for examining means of latent variables.

WLS was used to carry out analyses 1 through 3. For the multiple-group analyses (Analyses 2 and 3), Analysis 2 generated the less constrained model. Only the number of factors (the factor form) was constrained to be equal across the eight groups. Technically, in accord with recommendations of Byrne (1998), the plan in carrying out Analysis 1, 2, or 3 was to trim the measurement model if the practical measures of fit (i.e., in this research the RMSEA, NNFI, and the CFI) indicated an inadequate solution.[12]

---

[12] In this study, all practical measures of fit indicated adequate solutions for Analyses 1 and 2.

14

Analysis 1 involved the individual examination of the three-factor measurement model within each group. This was the baseline phase of the research (Byrne, 1998). Analysis 2 was carried out by constraining all eight groups simultaneously to have the same number of factors and same pattern of factor loadings per factor. Analysis 3 was carried out by creating a model in which the Λ_x (that is, the matrix of factor loadings) was constrained to be equal across the eight groups. This analysis involved the comparison of one nested model (that is a model with greater number of constraints and fewer estimated parameters) with a subsuming model with fewer constraints and a greater number of estimated parameters. In Analysis 3, the practical measures of fit (RMSEA, NNFI, and CFI) were examined to determine whether the constraint of equal factor loadings led to a relative decrement in practical fit. Little (1997) suggests that the practical measures of fit are adequate for assessing measurement-level equivalence which is embodied in Analysis 3.

In this regard, Little suggests that the NNFI can be compared as a way of indicating a loss of fit in a constrained (nested) model. He reported that "McGaw and Jöreskog (1971) concluded that an obtained difference in fit between a freely estimated solution and a constrained model of .022" in the NNFI was "negligible and opted for invariance on the basis of parsimony and this minimal difference in fit" (Little, 1997, p.58). Parsimony in SEM parlance refers to a solution with fewer numbers of parameters estimated—which occurs when certain parameters are held invariant over groups. Therefore, although no strict rules can be established as to an appropriate comparison of practical fit indices, there has been some indication that small differences in these indices are unimportant.

ML. ML analysis allows for a legitimate and accurate test of the fit of the model through the χ^2 fit index and the other practical fit measures in comparing latent mean structures. Because ML cannot be used with ordinal data of the type comprising attitude questionnaire data, the domain-representative method of creating parcels (sometimes called testlets) of items was employed (Kishton & Widaman, 1994). Each of the three scales was factor analyzed by means of principal axis method. A one-factor solution was examined. Parcels were formed by alternately selecting items with high, medium, and low loading values from the factor pattern matrix. For example, one parcel may consist of item one with the highest loading, item eight with the lowest loading, and item 10 with the midmost loading. The next parcel may consist of item five which of the remaining items had the highest loading, item seven which of the remaining items had the lowest loading, and item 3 which of the remaining items had the midmost loading. This process goes on until all items are included in a parcel. Parcel scores are computed by summing the item scores. These parcel scores can be treated as continuous variables and essentially are the indicator variables for the ML analysis.

In this research, the procedures just described were used to form three parcels for DCBTM and two parcels each for DTWM and RGSEP. Note that Analyses 2 and 3 were carried out through the application of the ML procedure on parcels. These analyses carried out with ML primarily serve as a means of proceeding to Analyses 4 though 7.

Analysis 4 was carried out by creating a model in which Λ_x and the diagonal elements of ϕ (that is, the variances of the latent constructs) were held invariant across groups. Because this involved a construct-level comparison (Little, 1997), the stricter statistical test employing $\Delta\chi^2$ was used to assess the equivalence of latent variances.

Analysis 5 required that Λ_x and the off-diagonal elements of ϕ are held invariant across the groups. Once again the $\Delta\chi^2$ was used as a test of equivalence of latent covariances.

Analysis 6 was carried out by creating a model in which the Λ_x (for parcels), the ϕ, and Θ_δ matrix (containing variances and covariances of indicator (parcels') residuals) were constrained to be equal across the groups. The $\Delta\chi^2$ and the practical fit measures were used here as well.

Finally, Analysis 7 was carried out. Analysis 7 involved constraining the τ_x (indicator intercepts) and the Λ_x (loadings) equal across groups. The κ (parameters representing latent means) values were fixed to zero in the first group and allowed to be freely estimated across the remaining groups. In this study, the African-American enlisted men comprised the first—in effect, the reference group. The group of African-American enlisted men is the largest minority group in the various services. Therefore, it makes practical and theoretical sense to place this group as the reference group. As a final part of Analysis 7, the means along with their standard errors and associated t-value (the ratio of the estimate to its standard error) are examined. Traditional values of significance (such as 1.96 for the .05 Type I error rate) can be used to determine whether a mean for a particular latent variable in a particular group is different from the mean of the reference group (which is fixed at a value of zero). However, in the case of many mean comparisons (for example, in this study, there are eight times three or 24 means), a correction for inflated Type I error was used. The overall Type I error rate was held at .05 with each individual Type I error rate set at .002.

Results

Analysis 1: Baseline Models

Appendix 2 presents the practical fit indices for the eight baseline models. RMSEA values from a high of .052 for White Women Officers to a low of .032 for White Male Officers. The minimum value for the NNFI and the CFI was .98. These data suggest that the practical fit of the three-factor models was acceptable. All Λ_x parameters in the models were statistically significantly different from zero.

Analysis 2: Constancy of Three-Factors

The difference between Analysis 1 and 2 is that Analysis 2 involved a multiple-group analysis and provided fit indices for the overall fit of the simultaneous solution. Because there were essentially no constraints, the overall chi-square test value is the sum of the individual chi-squares values. Goodness-of-Fit statistics for Analysis 2 are presented in Tables 3 and 4. Table 3 pertains to the first of the sample pairs and Table 4 pertains to the second of the sample pairs for all groups. All three practical fit indices indicate an acceptable fit in the data. Analysis 2 was replicated, in a sense, through the application of ML procedures. Tables 5 and 6 confirm that there is measurement equivalence across the eight groups and supported the item-level analyses carried out via WLS. In fact, with ML, the χ^2 values were statistically nonsignificant indicating strong evidence of constancy of the three-factor solution.

16

Analysis 3: Equivalence of Loadings

      Tables 3 and 4 also indicate that constraining the loadings (technically, the Λ_x matrix) to be equal across the eight groups did not appear to deteriorate the practical fit of the model. That is, the three practical fit indices were virtually identical in the first and second analyses. The statistically significant $\Delta\chi^2$ values in both samples ($\Delta\chi^2 = 425.88$, $p < .001$; and $\Delta\chi^2 = 496.48$, $p < .001$) were expected because with large samples, the $\Delta\chi^2$ test is extremely sensitive to even minor imperfections of fit in the data. However, the practical fit indices were of such a magnitude (close to optimal) that there appeared to be measurement invariance across all eight groups. Tables 5 and 6 indicate that the ML analyses also support the equivalence of the loadings. In fact, ML analyses (involving the use of parcels of items) showed that holding the

Table 3

Chi-Square Statistics and Goodness-of-Fit Indexes for the Measurement Models: Overall Invariance Across Eight Groups—Sample 1

Measurement model	Chi-square statistic			Goodness-of-fit indexes			Difference Statistics	
	df	χ^2	$p<$	NNFI	CFI	RMSEA	df	$\Delta\chi^2$
Constancy of 3-factor solution	1816	4999.12		.00	.99	.99	.042	--	--
Equivalent loadings	1956	5425.00		.00	.99	.99	.042	140	425.88*

Note. NNFI = Nonnormed fit index, CFI = Comparative fit index, and RMSEA = Root mean square error of approximation.

17

Table 4

Chi-Square Statistics and Goodness-of-Fit Indexes for the Measurement Models: Overall Invariance Across Eight Groups—WLS—Sample 2

Measurement model	df	χ^2	p<	NNFI	CFI	RMSEA	df	χ^2
	Chi-square statistic			Goodness-of-fit indexes			Difference Statistics	
Constancy of 3-factor solution	1816	5184.52	.00	.99	.99	.043	--	--
Equivalent loadings	1956	5681.00	.00	.99	.99	.044	140	496.48*

Note. NNFI = Nonnormed fit index, CFI = Comparative fit index, and RMSEA = Root mean square error of approximation.

Λ_x matrices in the eight groups resulted in nearly a perfect fit in terms of three practical fit indices (all of which indicated perfect fit) and a nonsignificant $\Delta\chi^2$ value (which indicated that the constraint did not significantly deteriorate the fit of the nested model vis a vis the subsuming model). In sum, there is strong evidence for measurement equivalence for the three latent variables across the eight sociocultural groups.

Analysis 4: Equivalence of Variances (in addition to Loadings) of Latent Variables

As indicated above, examination of the equivalence of latent variances is part of what Little (1997) calls a construct equivalence analysis. Therefore, the more conservative statistical fit index ($\Delta\chi^2$) with appropriate degrees of freedom is used to make inferences between the more and less restricted models. In the present study, this analysis involved the use of parcels and the ML analytic strategy. Tables 5 and 6 show that there is evidence in both samples for the lack of invariance of variance on the three constructs ($\Delta\chi^2$ (21) = 209.99, p <.05 in sample 1 and $\Delta\chi^2$ (21) = 177.55, p < .05, in sample 2). Stated another way, there is evidence of statistically significant differences in variability across the eight samples.

Analysis 5: Equivalence of Covariances (in addition to Loadings and Variances) Among Latent Constructs

Tables 5 and 6 indicate that covariances among latent constructs are noninvariant ($\Delta\chi^2$ (28) = 181.22, p < .05 in sample 1 and $\Delta\chi^2$ (28) = 176.94, p < .05).

Analysis 6: Equivalence of Indicator Measurement Error (in addition to Loadings and Latent Variances and Covariances)

Tables 5 and 6 indicate that the indicator error variances as well as error covariances are not invariant across groups ($\Delta\chi^2$ (49) =473.47, p <.05) in sample 1 and $\Delta\chi^2$ (49) = 435.38, p < .05 in sample 2).

Analysis 7: Equivalence of Construct Means (in addition to Loadings)

Finally, Tables 7 and 8 present the means with their standard errors. Because there are 24 separate t-tests presented in each table, it is appropriate to control for inflation of Type I error. This can be done by dividing the overall Type I error rate, .05, by 24 and because the differences may be direction, divided by 2. That is,

$$\alpha_{pc} = (.05/24)/2$$

where α_{pc} is the per comparison Type I error rate. Application of this formula resulted in a critical value of 3.00.

All comparisons presented in Tables 7 and 8 represented measurement-error-free comparisons. Findings indicate that African-American samples are substantially similar to one another (technically, not different from each other). In contrast, male and female White officers differ significantly from the African-American enlisted men on all three scales. The White groups are significantly more positive on all three scales. For the enlisted White samples, men are more positive on DCBTM and DTWM and women are more positive on DTWM. All results are replicated across pairs of random samples indicating that the results are quite free of sampling error.

Table 5

Chi-Square Statistics and Goodness-of-Fit Indexes for the Measurement Model—ML Analysis (Sample 1)

Measurement model	Chi-square statistic			Goodness-of-fit indexes			Difference Statistics	
	df	χ^2	$p<$	NNFI	CFI	RMSEA	df	$\Delta\chi^2$
Constancy of 3-factor solution	88	42.17	1.0	1.01	1.0	.00	--	--
Equivalent loadings	116	51.62	1.0	1.02	1.0	.00	28	9.45
Equivalent loadings & variances	137	261.61	.07	.98	.98	.045	21	209.99*
Equivalent loadings & phi	158	442.83	.00	.95	.95	.083	28	181.22*
Equivalent loadings, phi & theta	207	916.30	.00	.91	.88	.11	49	473.47*

Note. NNFI = Nonnormed fit index, CFI = Comparative fit index, and RMSEA = Root mean square error of approximation.

Table 6

Chi-Square Statistics and Goodness-of-Fit Indexes for the Measurement Model—ML Analysis (Sample2)

Measurement model	Chi-square statistic			Goodness-of-fit indexes			Difference Statistics	
	df	χ^2	$p<$	NNFI	CFI	RMSEA	df	χ^2
Constancy of 3-factor solution	88	45.74	1.0	1.01	1.0	.00	--	--
Equivalent loadings	116	60.85	1.0	1.01	1.0	.00	28	15.11
Equivalent loadings & variances	137	238.40	.00	.98	.98	.043	21	177.55*
Equivalent loadings & phi	158	415.34	.00	.96	.96	.080	28	176.94*
Equivalent loadings, phi & theta	207	850.72	.00	.92	.90	.11	49	435.38*

Note. NNFI = Nonnormed fit index, CFI = Comparative fit index, and RMSEA = Root mean square error of approximation.

Table 7

Group Means for Sample 1 (Parenthesized values are standard errors)

			DCBTM	DTWM	RGSEP
African Americans	Officers	Men	.83 (.34)	.17 (.40)	.54 (.24)
		Women	.23 (.35)	-.99 (.39)	.30 (.24)
	Enlisted	Men (reference)	0.00	0.00	0.00
		Women	-.02 (.34)	-.58 (.39)	.26 (.22)
White Americans	Officers	Men	3.63 (.27)*	5.18 (.33)*	1.20 (.20)*
		Women	3.06 (.29) *	3.37 (.36)*	1.37 (.19)*
	Enlisted	Men	2.36 (.30)*	3.90 (.35)*	0.00 (.24)
		Women	.90 (.34)	1.46 (.39)*	-.63 (.24)

NOTE: * p < .001 used to control for overall Type I error rate of .05

Table 8

Group Means for Sample 2 (Parenthesized values are standard errors)

			DCBTM	DTWM	RGSEP
African Americans	Officers	Men	.65 (.35)	.18 (.39)	.58 (.25)
		Women	.16 (.35)	-.88 (.39)	.60 (.24)
	Enlisted	Men (reference)	0.00	0.00	0.00
		Women	-.15 (.35)	-.67 (.39)	.35 (.23)
White Americans	Officers	Men	3.44 (.28)*	5.10 (.34)*	1.23 (.21)*
		Women	3.13 (.30)*	3.59 (.36)*	1.56 (.20)*
	Enlisted	Men	2.48 (.31)*	4.29 (.36)*	.31 (.24)
		Women	.72 (.35)	1.53 (.39)*	-.53 (.25)

NOTE: * $p < .001$ used to control for overall Type I error rate of .05

Discussion

The analysis of measurement and construct invariance provided a rich array of information, some straightforward and some complex.

Analyses pertaining to measurement equivalence provided information that for eight groups, the three scales (DCBTM, DTWM, and RGSEP) had equivalent meanings and could be used for further construct comparison. Recall that construct comparison involves not only the comparison of construct means but also the comparison of covariances among constructs, the variances of constructs, and the covariance of the constructs with external variables. In other words, measurement equivalence is a quality that provides the basis for extremely interesting and useful examinations of the variables in question. If measurement invariance had not been fully established, a series of post hoc analyses would have been required to determine whether measurement equivalence existed for some subset of the groups or to identify whether some subset of the constructs had complete measurement invariance across the groups. These post hoc analyses were not necessary in this case.

Analyses pertaining to construct equivalence provided a complex set of information. Analyses indicated that there were differences with regard to the variances of the latent constructs, covariances among the latent constructs, and means on the latent constructs. What does the finding of construct non-equivalence mean?

Noninvariance of Construct Variances and Covariances

Few organizational scientists and certainly very few practitioners in the field of training and organizational development address the issue of differences in variance. McIntyre (1997) examined variance differences in a much simpler design, which did not attend to the issue of measurement equivalence. He found evidence (crude in comparison to the findings here) of pairwise variance differences for five ethnic groups on all 12 MEOCS scales. McIntyre's 1997 recommendation is reinforced by the current study. It is important for organizational development professionals to examine whether variance differences exist and why they exist. This requires a focus-group type of research strategy where the members of the organization are consulted on as to their perceptions. In other words, the statistical research must be coupled with qualitative research to ascertain whether variance differences are a problem and, if so, which courses of action should be taken.

What about differences in covariances across groups? Here is where one might examine whether there is systematic gamma difference across the groups comprising the organization. The finding of differences in the covariances among constructs is an extremely interesting one and one which may inform training and organizational development specialists. Gamma difference implies that the perceptions, feelings, and beliefs reflected in the latent constructs are arrayed differently across groups. One group may believe that all of the issues are fairness issues and therefore have a high degree of relationship. Another group may see a greater distinction between, for example, RGSEP and DTWM, and hence this group would be characterized by a relatively lower degree of relationship between these two constructs than another group. The way constructs covary in a person's thinking in a sense reflects a schema or theory on which they behave, feel, perceive, and even respond to training. The importance of gamma difference, vis a

vis covariance among constructs, has important practical implications. As was the case with variance differences, a reasonable way of exploring covariance differences is through focus groups. The goal in such data gathering forums would be to understand the reasons for different covariance structures across groups. Once gathered, these data might lead to modification of training to better suit different groups. In fact, training modules on the gamma differences themselves may emanate from data such as these.

Mean Structures and Differences

A mean structures analysis was carried out to examine the difference between one reference group (enlisted African-American men) and seven other groups. Tables 7 and 8 clearly indicate similarities and differences in the levels of DCBTM, DTWM, and RGSEP in the groups compared. Because of the statistical power associated with the large sample sizes, the nonexistence of differences between the enlisted African-American males and the other African-American groups can be interpreted as similarities. Further, the fact that the mean structure analysis was effectively replicated in two random samples strengthens this interpretation. Differences were found between the reference group and White samples of men and women, officers and enlisted. Relatively large differences were found for the White officers (men and women) and White enlisted men. On the other hand, the enlisted women showed the least difference from the reference group with a statistically significant difference only on DTWM.

To administrators, the mean differences that were found may not be surprising. Studies in the past have found similar differences between African-American samples and the White samples. The use of a measurement-error-free approach with two random samples provides compelling confirmatory evidence for the military to continue to address racial tensions within the force. The message in this research is not that the state of affairs is extremely negative for African-Americans. Rather, it is that there are differences in attitudes that should be dealt with by military leaders striving for an optimal interracial relationships.

SEM allows for post hoc analyses to isolate the loci of differences that may account for the significant $\Delta\chi^2$ values. The set of purposes for this present research did not include exhaustive post hoc analyses. A broader scope was selected in order to give the reader a general understanding of the state of affairs as represented by three contructs comprising the MEOCS. Organizational development and training experts within the military would do well to continue with this line of research to gain a more indepth understanding of the nature of the construct nonequivalences that were found. Organizational scientists would do well to examine the theoretical meaning of construct nonequivalence and provide guidance on the phenomena that have been uncovered within this research.

Recommendations

The final goal of this research was to provide recommendations for the future. Therefore, the following key recommendations are made.

First, organizational researcher/practitioners involved in training design or organizational development should be aware of the measurement qualities of the instruments they use. They should understand the strengths and limitations of the instruments they use in organizational assessment. These strengths and weaknesses pertain to the construct validity of the instruments, the interpetability of the data that are collected by means of the instruments, and the logical actions that are suggested by these instruments. The present research along with the research carried out by McIntyre (1999) should serve to provide a set of guidelines for ensuring measurement quality.

Second, organizational scientists need to pay attention to the meaning of measurement and construct nonequivalence. Why do different sociocultural groups perceive measurement scales differently? Why are there differences in variances and covariances of latent constructs? Under what conditions do such differences make a practical difference for the training professional or the organizational development specialist? These are only some of the questions that organizational scientists should explore.

Third, organizational scientists and psychometricians need to work together to develop some theories on what constitutes the best way of developing organizational assessment instruments and attitude measures. Cultural differences associated with sociocultural groups affect the way instruments perform and the way they should be used. When many instruments are developed, developers fail to take into account from the beginning of the development how the different sociocultural groups comprising an organization will perceive the instrument. This lack of accounting may lead to measurement and construct nonequivalence. The result, especially of measurement nonequivalence, is a reduced opportunity to make sense of what organizational members and groups of organizational members experience.

A final recommendation concerns the MEOCS. The MEOCS—at least the three scales investigated in this study—showed measurement equivalence which fact is, at least in part, related to the intensive efforts used to develop the instrument (see Dansby and Landis, 1993). Users of the MEOCS should ensure that any comparisons that are made are made within the limitations of the data. The fact that certain construct nonequivalence exists is in itself an interesting finding which begs for understanding. Users should be aware that there is much more within the MEOCS database than simple mean differences between groups. Users of the MEOCS can take a lead in studying gamma differences between sociocultural groups and lead other organizational practitioners and researchers toward a deep understanding of organizational attitudes and climate.

References

Alwin, D.F. & Jackson, D.J. (1981). Application of simultaneous factor analysis to issues of factorial invariance. In D.J. Jackson & E.F. Borgatta (Eds.), Factor Analysis and Measurement. Beverly Hills, CA: Sage Publications Ltd.

Berndt, A. (1998). The Influence of Model Features on Goodness-of-Fit Indices: Choosing Indices to Evaluate Your Model. Unpublished Doctoral Dissertation, Old Dominion University.

Bollen, K. & Lennox, R. (1991). Conventional wisdom on measurement: a structural equation perspective. Psychological Bulletin, 110, 305-314.

Byrne, Barbara M. (1998). Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing or the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456-466.

Cole, D.A., Maxwell, S.E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. Psychological Bulletin, 114 (1), 174-184.

Dansby, M.R. (1993). Using and revising the MEOCS: An operational summary and methodology for updating and improving the Military Equal Opportunity Climate Survey. Defense Equal Opportunity Climate Survey Internal document.

Dansby, M. R. & Landis, D. (1991). Measuring equal opportunity climate in the military environment. International Journal of Intercultural Relations, 15, 389-405.

Golembiewski, R.T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. Journal of Applied Behavioral Science, 12, 133-157.

Hayduk, L.A. (1987). Structural equation modeling with LISREL. Baltimore, MD: The Johns Hopkins Press Ltd.

Kline, Rex B. (1998). Principles and Practice of Structural Equation Modeling. New York: The Guilford Press.

Jöreskog, K.G. & Sörbom, D. (1988). PRELIS—A program for multivariate data screening and data summarization. A preprocessor for LISREL. Second Edition. Chicago, Illinois: Scientific Software, Inc.

Jöreskog, K.G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. Quality and Quantity, 24, 387-404.

Jöreskog, K.G. & Aish, A.M. (1994). Structural Equation Modeling with Ordinal Variables. Chicago, Illinois: Scientific Software, Inc.

Jöreskog, K. G. & Sörbom, D. (1996). LISREL 8: Users Reference Guide, Chicago: Scientific Software International, Inc.

Kishton, J.M. & Widaman, K.F. (1994). Unidimensional versus domain representative parceling of questionnaire items: an empirical example. Educational and Psychological Measurement, 55 (3), 757-765.

Little, T.D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. Multivariate Behavioral Research, 32 (1), 53-76.

Little, T.D. (2000). On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. Journal of Cross-Cultural Psychology, March.

McGaw, B. & Jöreskog, K.G. (1971). Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. British Journal of Mathematical and Statistical Psychology, 24, 154-168.

McIntyre, R.M. (1998). Response styles and differences in variation of responses between demographic groups in the MEOCS data base. DEOMI Research Series Pamphlet 97-8. Patrick AFB, FL: Defense Equal Opportunity Management Institute.

Mortimer, J.T., Finch, M.D., & Kumka, D. (1982). Persistence and change in development: The multidimensional self-concept. In P.B. Baltes & O.G. Brim (Eds.), Lifespan development and behavior. (Vol. 4, pp. 263-313). San Diego: Academic Press.

Taris, T.W., Bok, I.A., & Meijer, Z.Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: a general approach. The Journal of Psychology, 132 (3), 301-316.

Appendix 1. Table of Common Measures of Fit Used in SEM (McIntyre, 1999)

Index	Abbreviation	Definition	Description	Prescribed Value	Advantages	Disadvantages
Minimum fit function Chi-Square (Likelihood Ration Test)	Chi-square	Traditional measure used to test the closeness of fit between the unrestricted sample covariance matrix S, and the restricted covariance matrix $\Sigma(\theta)$.	Tests the extent to which **all** residuals in $\Sigma - \Sigma(\theta) = 0$. The higher the probability associated with χ^2, the closer the fit.	No prescribed value. Lower values of nested models are preferred all other things being equal. Some recommend dividing by the df. Values greater than 2.0 (some say) or 3.0 (others say) are indicative of a poor fit.	It is the most commonly used fit index. If sample size is not too large, may provide information regarding overall model fit (Berndt, 1998).	It is highly affected by sample size. As sample size increases, so does the χ^2. In addition, there are a range of recommendations with regard to the cutoffs for the ration of this index to df.
Estimated Non-centrality parameter	NCP	Measure of the discrepancy between Σ and $\Sigma(\theta)$.	Natural measure of badness of fit. If the lower bound of the CI for the NCP encloses 0, then the model fits the data.	None.	Not much but seems to be an available measure of badness. Check the confidence interval.	Seems also to be affected by sample size.
Population Discrepancy Function Value	F_0	Estimated discrepancy between the fit that is estimated on the central Chi square distribution and the noncentral chi	Generally decreases as parameters increase.	None	Used to calculate other indices such as the RMSEA below	Affected by number of parameters

28

		square distribution			
Root Mean Square Error of Approximation	RMSEA	How well does the model with unknown but optimally chosen parameter values fit the population covariance matrix if it were available	Discrepancy is expressed per degree of freedom making it sensitive to the number of estimated parameters in the model	Less than .05: good fit >08 - .10: mediocre fit > .10: poor fit	Accepted as the most informative criteria in covariance structural modeling. Controls for sample size effects and complexity of model.	Subjectively determined "cutoff values"
Expected Cross-Validation Index	ECVI	Likelihood that the model cross-validates across similar-sized samples from the same population.	Discrepancy between the fitted covariance matrix in the analyzed sample and the expected covariance matrix that would be obtained in another sample of equivalent size	Compare with the index associated with the saturated model and the independence model. The model with the lowest index value has the greatest likelihood of replicating.	Useful in conceptualizing the replicability of the model	No set values really prescribed.

Chi-Square for Independence Model	Chi square— IM	Fit of the "null model"—that is, the model with no latent factors in CFA (E.G.)	Serves as a base against to compare to compare models. Expected to be large. Also used to compute NFI, NNFI, and CFI	Can be used to compare the obtained χ^2 value	Helps to understand the improvement in the hypothesized model.	Affected by sample size as are the other χ^2
Akaike's Information Criterion	AIC	Extent to which parameter estimates from original sample will cross-	Addresses the parsimony in the assessment of the model	Compare the AIC of the model to the independent and saturated	Takes parsimony as well as fit into account	No hard and fast rule for a cutoff. Comparative value is what counts.

			fit; statistical Goodness-of-Fit plus number of estimated parameters are taken into account	models.		
Consistent version of the Akaike Information Criterion	CAIC	Same as AIC but takes sample size into account	Addresses parsimony as in AIC as well as fit	Same as AIC	Same as AIC	Same as AIC
Root Mean Square Residual	RMR	Average residual value derived from the fitting of the variance-covariance matrix for the hypothesized value	Larger average values represent worse fit	No cutoffs generally recommend-ed.	Deals directly with the effect of model fit on the fitted covariance	Affected by the original metric of the variables and so is not very interpretable
Standardized RMR		Standardized RMR	Similar to RMR except that controls for the different scales of measurement makes interpretation much easier.	In a well-fitting model, the value will be .05 or less Byrne (1998).	As with RMR, Deals directly with the effect of model fit on the fitted covariance plus controls for scales of measure-ment.	No obvious disadvantages exist; however, the RMR seems not to be used very much
Goodness-of-Fit Index	GFI	Analogous to a squared multiple correlation in that it indicates the proportion of the observed covariances explained by the model implied covariances.	An absolute index of fit because it compares the hypothesized model to no model at all.	Closer to 1.0 is better.	Absolute	Affected by number of estimated parameters. Has fallen somewhat out of favor according to Berndt (19xx)

Adjusted Goodness-of-Fit Index	AGFI	Similar to GFI but corrected for the number of parameters much like an adjusted R-square value.	Takes parsimony into account and incorporates a "penalty" for inclusion of additional parameters	Closer to 1.0 is better.	Absolute and takes parsimony into account.	None discussed.
Parsimony Goodness-of-Fit Index	PGFI	Takes into account the complexity (number of estimated parameters) of the hypothesized model. Two logically interdependent pieces of information (GFI) and parsimony are represented by a single index.	Same	.50 are not unexpected. It appears as though greater than .50 should be a target.	The goal of assessing parsimony is its main advantage.	Little information on the targeted value may be a disadvantage.
Normed Fit Index	NFI	Proportion of improvement of the overall fit of the researcher's model relative to a null model.	It is an incremental fit index in that it is expressed as a percent improvement over the null model (an independence model)	Seems as though .90 or greater is recommended	Classic index for nearly a decade.	Does not control for model complexity. Adversely affected by sample size. (Berndt, 199x). Has fallen out of favor to some extent (Berndt, 19xx).

Non-Normed Fit Index	NNFI	Similar to NFI. Takes into account model complexity.	Sometimes referred to as the Tucker-Lewis Index.	Greater than .90 seems to be a common recommend-ation. Tucker and Lewis stated that it should be close to 1.0. Greater than .90 has been widely recommend-ed (Berndt, 1998).	Takes into account model complexity.	Can be difficult to interpret because it is not normed. Favors more "parsimonio us" or simpler models and penalizes more complex models.
Parsimony Normed Fit Index	PNFI	Equals the NFI multiplied by the parsimony ratio.	Takes the complexity of the model into account. Again, it is an incremental Goodness-of-Fit index.	Some say above .70 is expected.	Similar to other parsimony-indices.	
Comparative Fit Index	CFI	Like the NFI but takes sample size into account as well as df. Can range from 0 to 1.0.	Also is a comparative fit index comparing the hypothesize d to the independ-ence model	Closer to 1.0 is better. Greater than .90 has been widely recommend-ed (Berndt, 1998).	Another index that takes into account the sample size. Underestima tes fit less often than the NFI.	No clear indication of the relative value.
Incremental Fit Index	IFI	Same as the NFI except that degrees of freedom are taken into account	Same as NFI.	Close to 1.0.	Similar to above.	Similar to above.
Relative Fit Index	RFI	Algebraically equivalent to the CFI.	Same as CFI.	Close to 1.0.	Similar to above.	Similar to above.
Critical N	CN	Estimate of the	A test that is	>200	Provides	Doesn't say

		sample size that would be adequate to yield an adequate model fit for a χ^2 test.	independent of the effect of sample size.		information beyond that involved in incremental indices.	much on its own.

Appendix 2. Goodness of Fit Statistics for Eight Groups, Sample 1

Sample1: African-American Male Enlisted

```
              Goodness of Fit Statistics

                Degrees of Freedom = 227
    Minimum Fit Function Chi-Square = 579.78 (P = 0.0)
    Estimated Non-centrality Parameter (NCP) = 352.78

           Minimum Fit Function Value = 0.58
    Population Discrepancy Function Value (F0) = 0.35
Root Mean Square Error of Approximation (RMSEA) = 0.039

      Expected Cross-Validation Index (ECVI) = 0.68
            ECVI for saturated Model = 0.55
         ECVI for Independence Model = 25.82

Chi-Square for Independence Model with 253 Degrees of Freedom =
                       25752.99
           Independence AIC = 25798.99
                 Model AIC = 677.78
              Saturated AIC = 552.00
         Independence CAIC = 25934.87
               Model CAIC = 967.26
             Saturated CAIC = 2182.54

       Root Mean Square Residual (RMR) = 0.078
               Standardized RMR = 0.078
          Goodness of Fit Index (GFI) = 0.98
  Adjusted Goodness of Fit Index (AGFI) = 0.97
  Parsimony Goodness of Fit Index (PGFI) = 0.80

           Normed Fit Index (NFI) = 0.98
        Non-Normed Fit Index (NNFI) = 0.98
     Parsimony Normed Fit Index (PNFI) = 0.88
        Comparative Fit Index (CFI) = 0.99
        Incremental Fit Index (IFI) = 0.99
         Relative Fit Index (RFI) = 0.97

              Critical N (CN) = 482.58
```

Sample 1. African-American Male Officers

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 666.01 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 439.01

Minimum Fit Function Value = 0.67
Population Discrepancy Function Value (F0) = 0.44
Root Mean Square Error of Approximation (RMSEA) = 0.044

Expected Cross-Validation Index (ECVI) = 0.76
ECVI for saturated Model = 0.55
ECVI for Independence Model = 32.45

Chi-Square for Independence Model with 253 Degrees of Freedom =
32369.37
Independence AIC = 32415.37
Model AIC = 764.01
Saturated AIC = 552.00
Independence CAIC = 32551.25
Model CAIC = 1053.49
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.094
Standardized RMR = 0.094
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.97
Parsimony Goodness of Fit Index (PGFI) = 0.80

Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.98
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98

Critical N (CN) = 420.23

Sample 1. White Male Enlisted

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 564.92 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 337.92

Minimum Fit Function Value = 0.57
Population Discrepancy Function Value (F0) = 0.34
Root Mean Square Error of Approximation (RMSEA) = 0.039

Expected Cross-Validation Index (ECVI) = 0.66
ECVI for saturated Model = 0.55
ECVI for Independence Model = 29.52

Chi-Square for Independence Model with 253 Degrees of Freedom = 29445.70
Independence AIC = 29491.70
Model AIC = 662.92
Saturated AIC = 552.00
Independence CAIC = 29627.58
Model CAIC = 952.40
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.13
Standardized RMR = 0.13
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.98
Parsimony Goodness of Fit Index (PGFI) = 0.81

Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.99
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98

Critical N (CN) = 495.24

Sample 1.   White Male Officers

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 464.20 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 237.20

Minimum Fit Function Value = 0.46
Population Discrepancy Function Value (F0) = 0.24
Root Mean Square Error of Approximation (RMSEA) = 0.032

Expected Cross-Validation Index (ECVI) = 0.56
ECVI for saturated Model = 0.55
ECVI for Independence Model = 33:16

Chi-Square for Independence Model with 253 Degrees of Freedom = 33081.76
Independence AIC = 33127.76
Model AIC = 562.20
Saturated AIC = 552.00
Independence CAIC = 33263.64
Model CAIC = 851.68
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.11
Standardized RMR = 0.11
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.98
Parsimony Goodness of Fit Index (PGFI) = 0.81

Normed Fit Index (NFI) = 0.99
Non-Normed Fit Index (NNFI) = 0.99
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98

Critical N (CN) = 602.49

Sample 1. African-American Women Enlisted

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 577.49 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 350.49

Minimum Fit Function Value = 0.58
Population Discrepancy Function Value (F0) = 0.35
Root Mean Square Error of Approximation (RMSEA) = 0.039

Expected Cross-Validation Index (ECVI) = 0.68
ECVI for saturated Model = 0.55
ECVI for Independence Model = 28.10

Chi-Square for Independence Model with 253 Degrees of Freedom =
28024.04
Independence AIC = 28070.04
Model AIC = 675.49
Saturated AIC = 552.00
Independence CAIC = 28205.92
Model CAIC = 964.97
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.080
Standardized RMR = 0.080
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.97
Parsimony Goodness of Fit Index (PGFI) = 0.80

Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.99
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98

Critical N (CN) = 484.49

Sample 1. African-American Women Officers

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 718.61 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 491.61

Minimum Fit Function Value = 0.72
Population Discrepancy Function Value (F0) = 0.49
Root Mean Square Error of Approximation (RMSEA) = 0.047

Expected Cross-Validation Index (ECVI) = 0.82
ECVI for saturated Model = 0.55
ECVI for Independence Model = 28.00

Chi-Square for Independence Model with 253 Degrees of Freedom =
27930.65
Independence AIC = 27976.65
Model AIC = 816.61
Saturated AIC = 552.00
Independence CAIC = 28112.53
Model CAIC = 1106.09
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.11
Standardized RMR = 0.11
Goodness of Fit Index (GFI) = 0.97
Adjusted Goodness of Fit Index (AGFI) = 0.97
Parsimony Goodness of Fit Index (PGFI) = 0.80

Normed Fit Index (NFI) = 0.97
Non-Normed Fit Index (NNFI) = 0.98
Parsimony Normed Fit Index (PNFI) = 0.87
Comparative Fit Index (CFI) = 0.98
Incremental Fit Index (IFI) = 0.98
Relative Fit Index (RFI) = 0.97

Critical N (CN) = 389.54

Sample 1. White Women Enlisted

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 580.91 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 353.91

Minimum Fit Function Value = 0.58
Population Discrepancy Function Value (F0) = 0.35
Root Mean Square Error of Approximation (RMSEA) = 0.040

Expected Cross-Validation Index (ECVI) = 0.68
ECVI for saturated Model = 0.55
ECVI for Independence Model = 30.21

Chi-Square for Independence Model with 253 Degrees of Freedom =
30133.98
Independence AIC = 30179.98
Model AIC = 678.91
Saturated AIC = 552.00
Independence CAIC = 30315.86
Model CAIC = 968.39
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.10
Standardized RMR = 0.10
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.98
Parsimony Goodness of Fit Index (PGFI) = 0.81

Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.99
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.99
Incremental Fit Index (IFI) = 0.99
Relative Fit Index (RFI) = 0.98

Critical N (CN) = 481.64

Sample 1. White Women Officers

Goodness of Fit Statistics

Degrees of Freedom = 227
Minimum Fit Function Chi-Square = 847.20 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 620.20

Minimum Fit Function Value = 0.85
Population Discrepancy Function Value (F0) = 0.62
Root Mean Square Error of Approximation (RMSEA) = 0.052

Expected Cross-Validation Index (ECVI) = 0.95
ECVI for saturated Model = 0.55
ECVI for Independence Model = 34.94

Chi-Square for Independence Model with 253 Degrees of Freedom =
34856.69
Independence AIC = 34902.69
Model AIC = 945.20
Saturated AIC = 552.00
Independence CAIC = 35038.57
Model CAIC = 1234.68
Saturated CAIC = 2182.54

Root Mean Square Residual (RMR) = 0.11
Standardized RMR = 0.11
Goodness of Fit Index (GFI) = 0.97
Adjusted Goodness of Fit Index (AGFI) = 0.97
Parsimony Goodness of Fit Index (PGFI) = 0.80

Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 0.98
Parsimony Normed Fit Index (PNFI) = 0.88
Comparative Fit Index (CFI) = 0.98
Incremental Fit Index (IFI) = 0.98
Relative Fit Index (RFI) = 0.97

Critical N (CN) = 330.57